



一种室内人群时空聚集区域识别方法

王培晓¹ 吴升¹ 张恒才^{2,3} 陆锋^{2,3} 王宏恩⁴

1 福州大学数字中国研究院(福建),福建 福州,350002

2 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室,北京,100101

3 海西政务大数据应用协同创新中心,福建 福州,350002

4 山东科技大学测绘科学与工程学院,山东 青岛,266590

摘要:现有的人群聚集区域识别算法主要集中于室外空间,与室外空间相比,室内三维空间结构复杂,室内人群聚集更易导致安全事故的产生。提出了一种室内人群聚集区域识别方法——室内简化重构聚集方法(indoor simplification reconstruction cluster, IndoorSRC)。首先,设计了一种新型的室内时空凝聚层次聚类算法(indoor spatial-temporal agglomerative nesting, Indoor-STAGNES)识别室内用户停留点,简化室内用户移动轨迹;然后,构建了一种基于卡尔曼滤波的室内轨迹重构方法,实现了室内用户轨迹的对齐与重采样;最后,提出了一种室内时空密度聚类算法(indoor spatial-temporal ordering points to identify the clustering structure, Indoor-STOPTICS)发现室内三维时空人群聚集区域。采用真实室内轨迹数据进行实验分析,结果表明,与传统的室外识别方法相比,在运行时间差别不大的情况下,IndoorSRC识别的误识率可降低23.7%。

关键词:人群聚集区域识别;室内移动轨迹;卡尔曼滤波;Indoor-STAGNES;Indoor-STOPTICS

中图分类号:P208

文献标志码:A

室内空间是人类活动的主要空间,据统计,人类约有87%的时间都在室内空间移动^[1-2]。人群聚集是室内外应急安全事故的主要诱因之一,也是室内外安全监控的重点任务。与室外空间相比,室内三维空间结构更复杂^[3],人群聚集程度较高,较易发生拥挤、踩踏等安全事故。及时、快速、准确地识别室内空间的人群聚集区域对室内安全事故的提前预警与应急救援都具有重要意义^[4-5]。

近年来,随着室内定位技术不断发展,室内位置服务不断增多,移动终端设备不断普及,室内移动用户的轨迹数据呈爆发式增长,为室内相关研究提供了重要的数据源^[6-8],如文献[9-10]分别在传统指纹定位的基础上建立了隐式马尔科夫模型与置信区间模糊模型提高室内行人的定位精度;文献[11]结合室内用户轨迹及室内地标推断室内位置的语义信息;文献[12]利用室内WiFi定位数据推断用户的性别和年龄;文献[13]利用室内蓝牙定位数据预测用户下一时刻的位

置;文献[14]提出了一种数据驱动的方法寻找室内密集区域。上述与室内有关的研究主要集中于室内位置服务,并未识别室内人群聚集区域。目前,与人群聚集区域识别相关的研究主要集中于室外空间。例如文献[15]使用手机定位数据根据网格聚类探索室外人群聚集与消散模式;文献[16]利用GPS定位数据提出了一种基于密度的聚类算法识别路网级别的人群聚集区域,从而缓解交通拥堵现象。目前,现有的人群聚集区域识别算法集中于室外,相较于室外移动轨迹,室内移动轨迹存在质量差、易缺失、易停留等特点,传统的室外人群聚集区域识别算法难以发现室内人群聚集区域。例如,室内轨迹易缺失的特点将导致一个较大的室内人群聚集区域分裂为若干小区域,从而忽略了某些潜在的人群密集位置;室内轨迹易停留的特点将导致室内用户经常在某个特定区域产生较多轨迹点,从而形成“点聚集”而非“人聚集”区域,这在一定程度上降低了室内人群聚集区域识别的准确率。

收稿日期:2019-06-04

项目资助:国家自然科学基金(41771436,41701521,41771476);国家重点研发计划(2016YFB0502104,2017YFB0503500);数字福建建设项目(闽发改网数字函[2016]23号)。

第一作者:王培晓,硕士生,研究方向为地理信息服务与时空数据挖掘。peixiaowang@whu.edu.cn

通讯作者:张恒才,博士,助理研究员。zhanghc@reis.ac.cn

本文提出了一种新的室内人群聚集区域识别方法——室内简化重构聚集方法(indoor simplification reconstruction cluster, IndoorSRC)。该方法主要由 3 部分组成:室内移动轨迹简化、室内移动轨迹重构及室内人群聚集区域识别。其中,在室内移动轨迹简化阶段,本文提出了一种新的室内轨迹简化算法(indoor spatialtemporal agglomerative nesting, Indoor-STAGNES);在室内移动轨迹重构,卡尔曼滤波算法^[20-21]被用于简化轨迹的对齐与重采样;在室内人群聚集区域识别阶段,本文提出了一种新的室内时空密度聚类算法(indoor spatialtemporal ordering points to identify the clustering structure, Indoor-STOPTICS)。

1 室内简化重构聚集方法

室内人群聚集区域识别流程分为 4 个部分:(1)室内移动轨迹预处理,去除原始轨迹中异常、错误、无效的数据;(2)室内行人轨迹简化,提出 Indoor-STAGNES 算法简化室内用户移动轨迹,去除室内用户移动的停留信息;(3)室内行人轨迹重构,根据室内轨迹采样间隔信息及卡尔曼滤波算法重构室内用户,简化轨迹点;(4)室内人群聚集区域提取,提出 Indoor-STOPTICS 算法为重构轨迹点生成增广簇排序图,依据增广簇排序图辅助提取室内人群高密度区域。相关定义见图 1。

定义 1 原始轨迹点 $p = \langle u_{id}, t, x, y, f \rangle$, p 是移动终端采集的记录信息, u_{id} 是用户唯一标

识, t 代表记录信息采集的时间, (x, y) 代表该用户在 t 时刻的坐标信息, f 表示该用户在 t 时刻所处楼层。

定义 2 原始轨迹 $T = \{p_i\}_{i=1}^n$, T 为某用户按时间顺序连接的原始轨迹点序列(见图 1(b)), n 为某用户原始轨迹点记录个数。

定义 3 简化轨迹点 p^{sim} , 若干时间连续且位置移动较小的待简化轨迹点团簇经简化得到 $p^{sim} = \langle u_{id}, t^{sim}, x^{sim}, y^{sim}, f \rangle$, 如图 1(c) 所示, t^{sim} 为待简化团簇平均记录时间, (x^{sim}, y^{sim}, f) 代表简化团簇的位置, (x^{sim}, y^{sim}) 为待简化团簇的中心坐标, f 为待简化团簇所在楼层。

定义 4 简化轨迹 $T^{sim} = \{p_i^{sim}\}_{i=1}^k$, T^{sim} 为某用户按时间顺序连接的简化轨迹点序列, k 为某用户简化轨迹点记录个数。

定义 5 重构轨迹点 $p^{rec} = \langle u_{id}, t, \hat{x}, \hat{y}, f \rangle$, 简化轨迹点经重构后得到重构轨迹点 p^{rec} , 如图 1(d) 所示, t 表示重构轨迹点 p^{rec} 的记录时间, (\hat{x}, \hat{y}) 表示该用户在 t 时刻重构后的坐标信息, f 表示该用户在 t 时刻所处楼层。

定义 6 重构轨迹 $T^{rec} = \{p_i^{rec}\}_{i=1}^m$, T^{rec} 为某用户按时间顺序连接的重构轨迹点序列, m 为某用户重构轨迹点记录个数。

定义 7 重构轨迹点集合 $D = \{p_i^{rec}\}_{i=1}^N$, D 为用于识别室内人群聚集区域的所有轨迹点组成的集合, N 为群体用户所有重构轨迹点的个数。

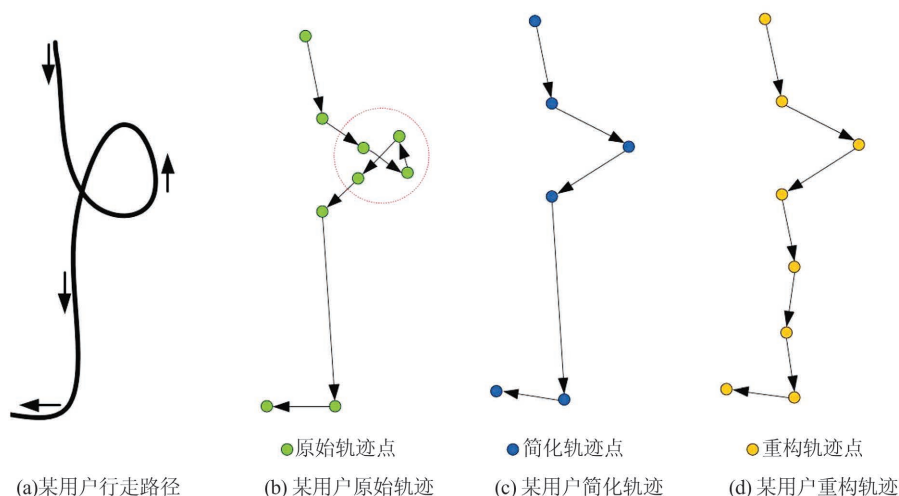


图 1 相关定义

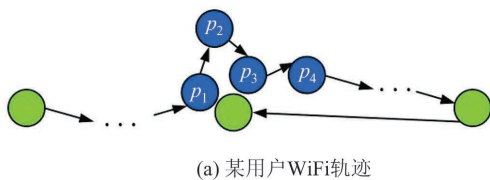
Fig.1 Related Definitions

1.1 室内行人轨迹简化

室内定位轨迹时间间隔存在不均质特性,尤

其是当移动目标在特定地方停留一定时间时,将导致移动终端在有限区域内记录较多的轨迹点,

从而形成轨迹点簇团,若直接使用原始轨迹识别人群聚集区域,往往得到轨迹点聚集区域而非人群聚集区域。因此,本文提出 Indoor-STAGNES 算法对用户轨迹进行简化,去除轨迹点停留信息。Indoor-STAGNES 算法是传统凝聚层次聚类 (agglomerative nesting, AGNES) 算法的改进算法。该算法在 AGNES 算法的基础上引入了时间窗口 Δt 及楼层约束,通过迭代的方式不断把同楼层时空邻近的轨迹点(簇集)合并,最终将原始轨迹 $T = \{p_i\}_{i=1}^n$ 划分为 k 个不相交顺序簇集 $\{C_1, C_2 \dots C_k\}$,簇集 C_i 内的原始轨迹点集经简化得到 p_i^{sim} , k 个簇集得到 k 个简化轨迹点,即 $T^{\text{sim}} = \{p_i^{\text{sim}}\}_{i=1}^k$,其中 $k \leq n$ 。如图 2 所示,簇集 $C_i(p_1, p_2, p_3, p_4)$ 迭代为一个簇集,最终将简化为



(a) 某用户WiFi轨迹

一个简化轨迹点 p_i^{sim} ,迭代过程中簇集之间空间距离的计算公式为:

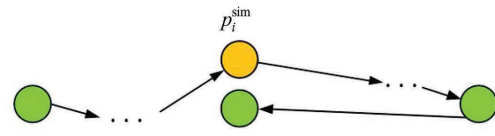
$$f_{\text{spatial_dist}}(C_i, C_j) = \|\bar{p}_i - \bar{p}_j\|_2, \bar{p}_i = \frac{1}{\text{card}(C_i)} \sum_{p_i \in C_i} p_i \quad (1)$$

式中, $f_{\text{spatial_dist}}$ 函数计算簇集 C_i 与 C_j 的空间距离; $\text{card}(C_i)$ 表示簇集 C_i 包含的轨迹点个数; \bar{p}_i, \bar{p}_j 表示簇集 C_i, C_j 的均值坐标。

簇集 C_i 与 C_j 的时间距离计算公式为:

$$f_{\text{time_dist}}(C_i, C_j) = |C_i \cdot t_{\text{ave}} - C_j \cdot t_{\text{ave}}|, C_i \cdot t_{\text{ave}} = \frac{1}{\text{card}(C_i)} \sum_{p_i \in C_i} p_i \cdot t \quad (2)$$

式中, $f_{\text{time_dist}}$ 函数计算簇集 C_i 与 C_j 的时间距离; $C_i \cdot t_{\text{ave}}, C_j \cdot t_{\text{ave}}$ 分别表示簇集 C_i, C_j 中轨迹点的平均记录时间。



(b) 原始轨迹点简化结果

图 2 Indoor-STAGNES 算法简化室内移动轨迹

Fig.2 Indoor Trajectory Simplification with Indoor-STAGNES

Indoor-STAGNES 算法的具体流程如下:

1) 输入时间连续的室内用户轨迹 $T = \{p_i\}_{i=1}^n$,并将每一个轨迹点初始化为一个簇集。

2) 初始化簇集之间的空间距离矩阵 M^S 及时间距离矩阵 M^T , M_{ij}^S 表示簇集 C_i 和簇集 C_j 之间的空间距离, M_{ij}^T 表示簇集 C_i 和簇集 C_j 之间的时间距离,若 C_i 与 C_j 不属于同一个楼层,则 M_{ij}^S 与 M_{ij}^T 为无穷大。

3) 寻找时间约束 Δt 下距离矩阵 M^S 中的最小值 d_{\min} ,如果 d_{\min} 小于距离阈值 d_{threh} ,则将最近两个簇集合并,同时更新空间距离矩阵 M^S 与时间距离矩阵 M^T ,再次寻找 d_{\min} ;如果 d_{\min} 大于距离阈值 d_{threh} ,则得到最终的簇集,跳转至步骤 4)。

4) 按照时间顺序依次将簇集 $\{C_1, C_2 \dots C_k\}$ 简化为 $\{p_1^{\text{sim}}, p_2^{\text{sim}} \dots p_k^{\text{sim}}\}$ 。

1.2 室内行人轨迹重构

经简化的轨迹 T^{sim} 大体反映了用户的移动骨架,但仍不适合识别室内人群聚集区域。简化轨迹 T^{sim} 中包含较多的缺失轨迹点,若直接采用简化轨迹 T^{sim} 识别室内人群聚集区域,在一定程度上将会影响室内人群聚集区域识别的准确率。

因此,室内移动轨迹重构是室内人群聚集区域识别的重要环节之一。为补齐简化轨迹中缺失的轨迹点,本文采用卡尔曼滤波算法重构室内用户简化轨迹。卡尔曼滤波是一种线性最优估计算法,综合考虑测量数据与物理运动模型,通过迭代的方式依次估计某用户每一个时刻的最优位置,即重构轨迹点 p^{rec} 。使用卡尔曼滤波算法重构室内移动轨迹主要分为如下两个阶段:

1) 缺失轨迹点确定。依据简化轨迹 T^{sim} 的采样间隔确定简化轨迹中缺失轨迹点位置及个数,如图 3 所示,将简化轨迹 T^{sim} 上相邻轨迹点采样间隔超两倍平均采样间隔的轨迹区间视为存在缺失轨迹区间,当缺失轨迹区间采样间隔小于采样间隔第 95 百分位数时,缺失轨迹点需补齐。缺失轨迹点个数与时间信息的计算公式为:

$$c = f\left(\frac{m_{\text{interval}}}{a_{\text{interval}}}\right) + 1 \quad (3)$$

$$p_{i+j}^{\text{rec}} \cdot t = p_i^{\text{rec}} \cdot t + ja_{\text{interval}}, 1 \leq j \leq c \quad (4)$$

式中, c 表示缺失轨迹点个数; m_{interval} 表示缺失轨迹区间的采样间隔; a_{interval} 表示简化轨迹点的平均采样间隔; $f(x)$ 表示向下取整函数; $p_{i+j}^{\text{rec}} \cdot t$ 表示缺

失区间第 j 个缺失轨迹点的时间信息,如 $p_{i+1}^{\text{rec}}.t$ 表示该缺失区间第 1 个缺失轨迹点的时间信息。

2) 迭代重构移动轨迹。采用卡尔曼滤波算法迭代求解每一时刻的重构轨迹点的位置,主要分为位置预测与位置更新两个阶段:在位置预测

阶段,依据上一时刻的最优位置采用物理运动模型预测下一时刻的预测位置;在位置更新阶段,利用当前时刻的测量数据及误差修正当前时刻的预测位置得到当前时刻的最优位置,最优位置即为当前时刻重构轨迹点的位置。

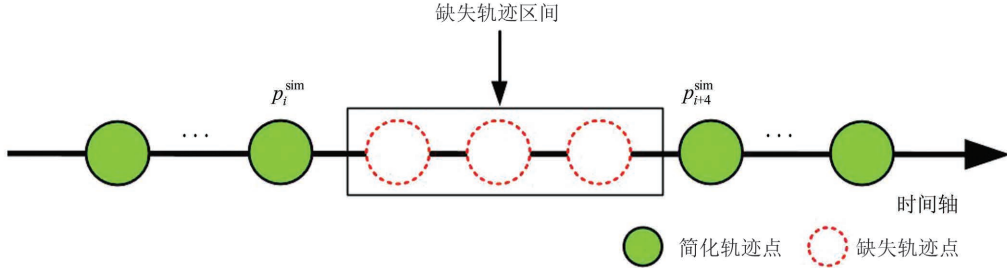


图 3 包含缺失轨迹点的简化轨迹

Fig.3 Simplified Trajectory Containing Missing Points

1.3 室内人群聚集区域提取

经重构后的重构轨迹点 p^{rec} 较精确地反映了用户位置随时间的移动,将所有用户的重构轨迹点 p^{rec} 组合在一起即可分析室内群体用户位置随时间的变化,从而识别室内人群聚集区域。本文将重构轨迹点集合 D (定义 7) 中的高密度区域视为室内人群聚集区域,为此提出了基于密度的室内时空聚类算法 Indoor-STOPTICS。

定义 8 时空邻域,对于重构轨迹点 $p_i^{\text{rec}} \in D$,其时空邻域表示以 p_i^{rec} 为中心,以 ϵ_1 为空间半径、以 ϵ_2 为时间窗口所围成的圆柱体,如图 4 所示, $N_{\epsilon_1, \epsilon_2}(p_i^{\text{rec}})$ 为圆柱体内部所包含的轨迹点子集合,计算公式为:

$$N_{\epsilon_1, \epsilon_2}(p_i^{\text{rec}}) = \{ p_j^{\text{rec}} \in D \mid s_d(p_j^{\text{rec}}, p_i^{\text{rec}}) \leq \epsilon_1 \text{ 且 } t_d(p_j^{\text{rec}}, p_i^{\text{rec}}) \leq \epsilon_2 \text{ 且 } p_j^{\text{rec}}.f = p_i^{\text{rec}}.f \} \quad (5)$$

式中, s_d 用于度量两轨迹点的空间距离; t_d 用于度量两轨迹点的时间距离; ϵ_1 为空间半径阈值; ϵ_2 为时间窗口阈值。轨迹点子集合个数记为 $|N_{\epsilon_1, \epsilon_2}(p_i^{\text{rec}})|$ 。

Indoor-STOPTICS 算法是时空密度聚类 (spatial-temporal ordering points to identify the clustering structure, ST-OPTICS)^[19-20] 算法的改进算法,由于室内轨迹具有三维特性,Indoor-STOPTICS 算法在 ST-OPTICS 的基础上添加了楼层约束(定义 8)。该算法用于寻找时空约束下同楼层最大密度相连的轨迹点集合。与传统基于密度的时空聚类算法不同,Indoor-STOPTICS 算法并不显式地生成数据聚类,而是为聚类分析生成一个有序的对象列表,辅助识别集合 D 中的高密度区域(室内人群聚集区域)。Indoor-

STOPTICS 算法的具体流程如下:

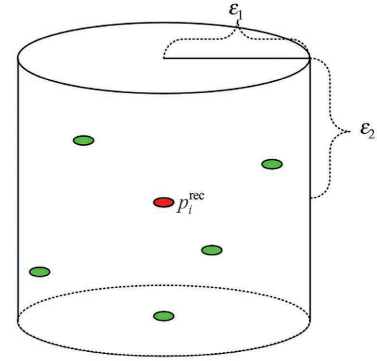


图 4 时空邻域

Fig.4 Spatiotemporal Neighborhood

1) 输入重构轨迹点集合 $D = \{p_i^{\text{rec}}\}_{i=1}^N$, 将每一个重构轨迹点的可达距离设置为空值,并初始化有序轨迹点列表 $L = \emptyset$ 。

2) 从集合 D 中寻找未处理的轨迹点 p_i^{rec} , 若所有轨迹点均已处理,跳转至步骤 4), 否则跳转至步骤 3)。

3) 获取 p_i^{rec} 时间约束下同楼层的邻居点集合 $N_{\epsilon_1, \epsilon_2}(p_i^{\text{rec}})$, 并将 p_i^{rec} 设置为已处理, 添加到列表 L , 若 $|N_{\epsilon_1, \epsilon_2}(p_i^{\text{rec}})|$ 大于最小点个数 M_{threh} , 则将 $N_{\epsilon_1, \epsilon_2}(p_i^{\text{rec}})$ 中的轨迹点根据可达距离从小到大依次加入到种子集合 L_{seeds} 中。(1) 判断 L_{seeds} 是否为 \emptyset , 若为 \emptyset , 跳转到步骤 2), 否则, 从 L_{seeds} 中取出第一个轨迹点 q , 获得其邻居点集合 $N_{\epsilon_1, \epsilon_2}(q)$, 并将 q 设置为已处理, 添加到列表 L ; (2) 判断 $N_{\epsilon_1, \epsilon_2}(q)$ 是否大于 M_{threh} , 若小于, 跳转到步骤 1), 否则将 $N_{\epsilon_1, \epsilon_2}(q)$ 中的轨迹点根据可达距离从小到大

大依次加入到种子集合 L_{seeds} 中。

4) 获得集合 D 的有序轨迹点列表 L 。如图 5 所示, 重构轨迹点集合 $D = \{p_i^{rec}\}_{i=1}^N$, 经 Indoor-STOPTICS 算法得到有序轨迹点列表 $L = \{p_j^{rec}\}_{j=1}^N$, 以 L 的下标索引 j 为横轴, p_j^{rec} 的可达距离为纵轴可绘制集合 D 的增广簇排序图。从图 5 中可得辅助信息如下:

- 1) 当 Indoor-STOPTICS 算法空间半径 $\epsilon_1 = r_1$ 时, 可从集合 D 中识别两个簇集 $A_{cluster}$ 和 $B_{cluster}$ 。
- 2) 当 Indoor-STOPTICS 算法空间半径 $\epsilon_1 =$

r_2 时, 簇集 $A_{cluster}$ 分裂成 3 个小簇, 即 $A_{cluster}^1$ 、 $A_{cluster}^2$ 和 $A_{cluster}^3$, 此时可从集合 D 中识别 4 个簇集。

3) 当 Indoor-STOPTICS 算法空间半径 $\epsilon_1 = r_2$ 时, 每一个簇集包含的轨迹点可通过对应横轴索引序列得到, 如簇集 $A_{cluster}^1$ 所对应的横轴索引序列为 X_{idxArr} , 簇集 $A_{cluster}^1$ 所包含的轨迹点为 $\{L[i]\}_{i \in X_{idxArr}}$ 。

4) 当 Indoor-STOPTICS 算法空间半径 $\epsilon_1 = r_2$ 时, 簇集的大小及密集程度可定量表示。如簇集 $B_{cluster}$ 的大小可由 w 表示, w 越宽, 簇集越大; $B_{cluster}$ 的密集程度可由 h 表示, h 越低, 簇集越紧凑。

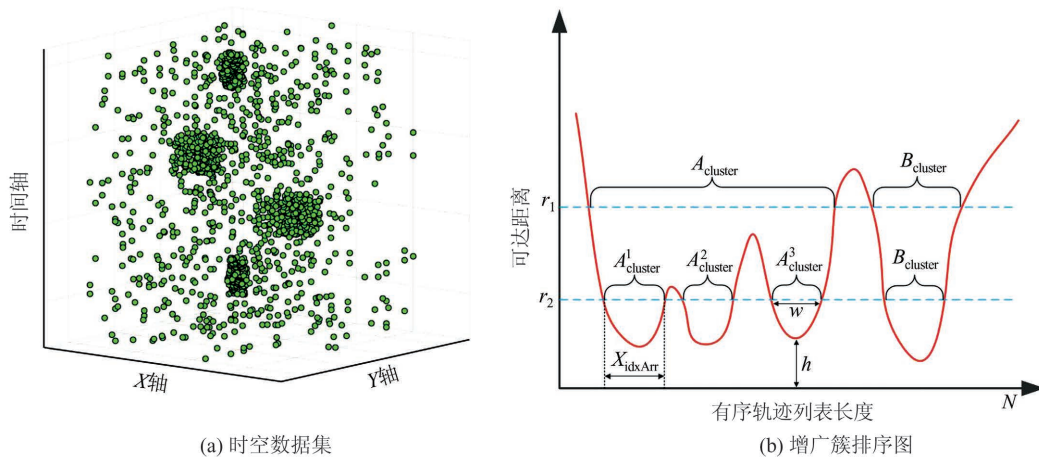


图 5 时空数据集及其增广簇排序图

Fig.5 Spatiotemporal Dataset and Illustration of the Cluster-Ordering

2 室内人群聚集区域识别结果与分析

2.1 实验数据

实验数据来源于济南市某商场 1 d 的移动用户 WiFi 定位数据, 数据覆盖商场 8 个楼层, 共采集室内用户移动轨迹约 200 万条, 总轨迹点逾 3 000 万个, 采样间隔为 1~5 s 的轨迹点占比 70% 以上, 数据定位精度约为 3 m, 数据字段包括用户唯一标识 (identification, ID)、记录上传时间、用户所在位置 (投影坐标 X、Y 及所在楼层 ID), 如表 1 所示。

2.2 算法评价指标及对比实验选择

本文以人群密度、点群密度、误识率及算法运行时间作为室内人群聚集区域识别的定量评价指标。其中人群密度 C_d 、点群密度 P_d 和误识率 F 的计算公式分别为:

$$C_d = \sum_i^n \frac{N_i^{crowdNum}}{V_i} \times \frac{\Delta t}{n} \quad (6)$$

表 1 室内用户移动轨迹实例

Tab.1 Indoor User Trajectory Sample

用户 ID	上传时间	X/m	Y/m	楼层 ID
8026BC4***	2017-12-31T10:00:01	130219***	43903***	2
8026BC4***	2017-12-31T10:00:03	130219***	43903***	2
8026BC4***	2017-12-31T10:00:04	130219***	43903***	2
⋮	⋮	⋮	⋮	⋮
8026BC4***	2017-12-31T14:59:54	130219***	43903***	4

$$P_d = \sum_i^n \frac{N_i^{pointNum}}{V_i} \times \frac{\Delta t}{n} \quad (7)$$

$$F = \frac{G_{error}}{G_{algorithm}} \quad (8)$$

式中, n 表示算法识别的簇集个数; V_i 表示某簇集的体积, 即三维点集的凸包体积; $N_i^{crowdNum}$ 表示某簇集中包含用户的数量; Δt 表示时间步长, 即 C_d 表示在时间 Δt 内簇集单位体积中累积的用户数量; $N_i^{pointNum}$ 表示某簇集中包含轨迹点的数量, 即 P_d 表示在时间 Δt 内簇集单位体积中累积的轨迹点数量; $G_{algorithm}$ 表示算法识别的人群聚集区域个

数; G_{error} 表示算法判断错误的人群聚集区域个数。当聚集区域内的用户数量小于数量阈值或聚集区域的聚集时间小于时间阈值时,区域被标识为识别错误。

本文采用 ST-OPTICS 和 IndoorSRC 算法进行对比,探讨两种算法在不同用户数量下误识率、人群密度、点群密度及算法运行时间的变化情况。ST-OPTICS 算法利用预处理后轨迹的点集直接识别室内人群聚集区域,IndoorSRC 算法对预处理后点集进行简化和重构,采用 Indoor-STOPTICS 算法识别室内人群聚集区域。

2.3 识别结果与分析

本文采用定性和定量两种方式分析 IndoorSRC 算法与 ST-OPTICS 算法的识别结果,定性方式主要采用图形的方式直观展示两种方法在

某楼层的识别结果,定量的方式主要采用相关指标评价两种方法的性能。

以 11 时至 16 时作为研究时段,IndoorSRC 方法在某楼层的识别结果如图 6 所示,重构轨迹点集合 D 在不同区域、不同时间呈现明显的聚集模式,绘制集合 D 的增广簇排序图,可具体识别高密度区域存在的位置及时间。以空间半径 $\epsilon_1=6$ m 为例,集合 D 将存在 8 个高密度区域,进一步分析每一个簇集大小与密集程度,最终得到 5 个室内人群聚集区域。5 个室内人群聚集区域位置及时间特征如表 2 所示。聚集区域多集中于中午时段且主要位于餐饮地区,如食尚家属于商场内部的小吃区,在中午时段形成 3 个时间较短的人群聚集区,渔家灯火和中餐厅属于商场内部的饭店,各形成一个时间较长的人群聚集区。

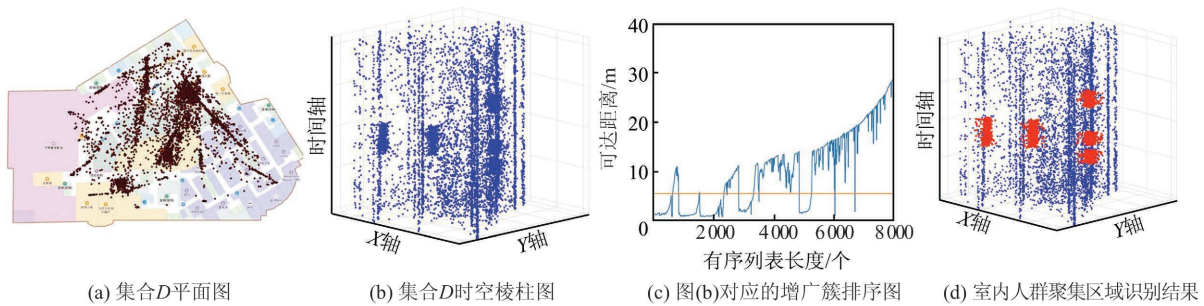


图 6 室内人群聚集区域识别结果

Fig.6 Indoor Crowd Gathering Area Identification Results

表 2 室内人群聚集区域位置时间特征

Tab.2 Indoor Crowd Gathering Area Location and Time Characteristics

聚集区域	聚集位置	聚集时间
聚集区域 1	食尚家	11:31:15—11:56:23
聚集区域 2	食尚家	12:07:24—12:34:09
聚集区域 3	中餐厅	12:16:17—13:32:48
聚集区域 4	渔家灯火	12:20:03—13:24:38
聚集区域 5	食尚家	14:14:13—14:17:29

ST-OPTICS 算法的结果如图 7 所示,与 In-

doorSRC 算法识别结果相比:(1)ST-OPTICS 算法虽然在不同区域、不同时间发现了较明显的聚集模式,但该聚集模式多为“点聚集”而非“人群聚集”,即由于用户购物目的性导致的停留点集合;(2)ST-OPTICS 算法识别的簇集中存在较多相邻较近的小簇集,原因是室内轨迹数据质量较差,存在较多缺失数据,直接采用轨迹点信息识别室内人群聚集区域将导致本是一个的簇集分裂成较多个小簇集。

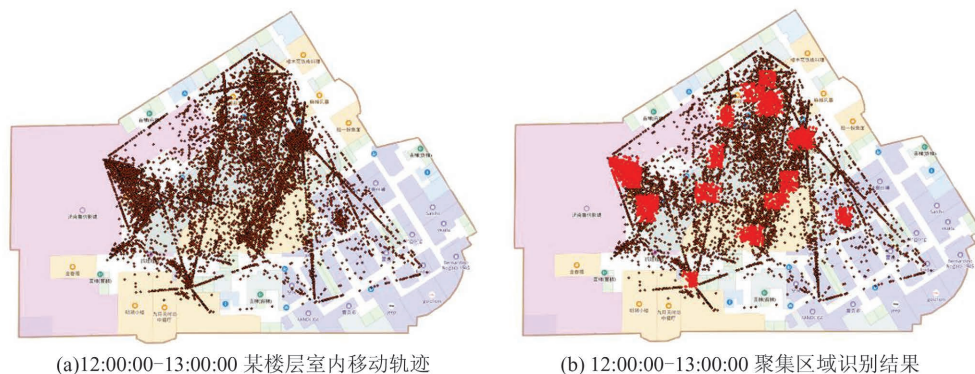


图 7 ST-OPTICS 算法识别结果

Fig.7 Recognition Result of ST-OPTICS Algorithm

IndoorSRC算法与ST-OPTICS算法人群密度、点群密度、误识率和运行时间的对比见图8。由图8可知:(1)IndoorSRC算法识别的人群密度远高于ST-OPTICS算法识别的人群密度,而ST-OPTICS算法识别的点群密度略高于IndoorSRC算法的点群密度,原因是ST-OPTICS算法识别的簇集较多为“点聚集”而非“人聚集”区域。(2)当用户量较少时,IndoorSRC算法的运行时间略高于ST-OPTICS算法的运行时间,原

因是IndoorSRC算法对预处理后的轨迹做了简化与重构,使得算法的运行时间增加,但随着用户量的增加,预处理后轨迹点数量远大于重构轨迹点数量,使得ST-OPTICS算法的运行时间超过IndoorSRC算法的运行时间。(3)ST-OPTICS算法的误识率随用户数量的增加缓慢提升,原因是随着用户数量的增多,用户的停留区域也逐渐增多,在一定程度上提高了ST-OPTICS算法的误识率。

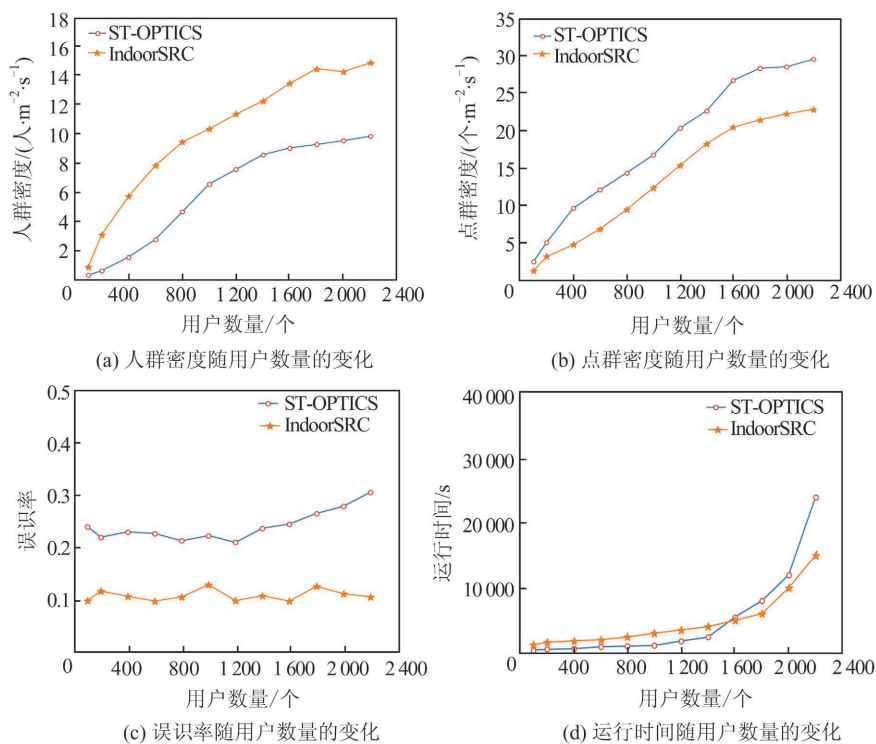


图8 IndoorSRC和ST-OPTICS定量对比结果

Fig.8 Quantitative Comparison of IndoorSRC and ST-OPTICS

4) IndoorSRC算法的误识率较稳定,且远低于ST-OPTICS算法的误识率,误识率最高可降低23.7%,原因是IndoorSRC算法对室内用户轨迹做了简化与重构,降低了用户停留区域和缺失轨迹点对识别结果的影响。

3 结语

针对传统的室外人群聚集区域识别算法难以应用到室内空间的问题,本文提出了一种室内的人群聚集区域识别框架 IndoorSRC。首先,提出了 Indoor-STAGNES 算法简化室内用户移动轨迹,去除用户停留信息;接着,构建了一种基于卡尔曼滤波的室内轨迹重构方法,补齐缺失轨迹点信息;最后提出了 Indoor-STOPTICS 算法发现室内三维时空人群聚集区域。实验结果表明,

与传统算法运行时间差别不大的情况下,IndoorSRC算法能识别的准确率可提高23.7%。本文能够为室内人群聚集区域识别方法提供有效补充,为室内安全事故的提前预警与应急救援提供技术支持。

室内人群集聚仅是室内拥挤、踩踏等安全事故产生的必要条件,除此之外,用户移动方向同样在很大程度上影响了室内安全事故的产生,例如,当人群集聚且同方向行走时往往不会发生安全事故,只有当人群集聚且行走方向无序时将较大可能导致安全事故的发生。因此,未来将在 IndoorSRC 方法的基础上引入方向等其他约束,提高识别的准确率,从而辅助室内安全事故的提前预警与应急救援。

致谢:感谢上海图聚智能科技有限公司提供室内定位轨迹实验数据。

参 考 文 献

- [1] Klepeis N E, Nelson W C, Ott W R, et al. The National Human Activity Pattern Survey (NHAPS): A Resource for Assessing Exposure to Environmental Pollutants[J]. *Journal of Exposure Analysis and Environmental Epidemiology*, 2001, 11(3):231-252
- [2] Zhou Chenghu. Prospects on Pan-Spatial Information System[J]. *Progress in Geography*, 2015, 34(2):129-131(周成虎. 全空间地理信息系统展望[J]. 地理科学进展, 2015, 34(2):129-131)
- [3] Wu Enchao, Zhang Hengcai, Wu Sheng. Automatic Generation Method of Indoor and Outdoor Integrated Navigation Network Based on Medial Axis Transform Algorithm[J]. *Journal of Geo-Information Science*, 2018, 20(6): 730-737(武恩超, 张恒才, 吴升. 基于中轴变换算法的室内外一体化导航路网自动生成方法[J]. 地球信息科学学报, 2018, 20(6):730-737)
- [4] Ren Changxing, Wu Zongzhi, Liu Mao. Analysis of Accident Due to Swarm and Jostlement at Urban Public Place[J]. *China Safety Science Journal*, 2005, 15(12): 102-107(任常兴, 吴宗之, 刘茂. 城市公共场所人群拥挤踩踏事故分析[J]. 中国安全科学学报, 2005, 15(12): 102-107)
- [5] Zhang Qingsong, Liu Jinlan, Zhao Guomin. Discussion on Cause Mechanism of Crowd Crushing and Trampling Accident in Large-Scale Public Places[J]. *Journal of Natural Disasters*, 2009, 18(6):81-86(张青松, 刘金兰, 赵国敏. 大型公共场所人群拥挤踩踏事故机理初探[J]. 自然灾害学报, 2009, 18(6):81-86)
- [6] Wang P, Wu S, Zhang H, et al. Indoor Location Prediction Method for Shopping Malls Based on Location Sequence Similarity[J]. *ISPRS International Journal of Geo-Information*, 2019, 8(11): 1-17
- [7] Wang P, Wang H, Zhang H, et al. A Hybrid Markov and LSTM Model for Indoor Location Prediction[J]. *IEEE Access*, 2019, 7(1): 185 928-185 940
- [8] Park J K, Woo T G, Kim M, et al. Hadamard Matrix Design for a Low-Cost Indoor Positioning System in Visible Light Communication[J]. *IEEE Photonics Journal*, 2017, 9(2):1-10
- [9] Ayong Y, Jianfei S, Li X, et al. Local HMM for Indoor Positioning Based on Fingerprinting and Displacement Ranging [J]. *IET Communications*, 2018, 12(10):1 163-1 170
- [10] Simon T, Dejan D, Igor S. Confidence-Interval Fuzzy Model-Based Indoor Localization[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(3): 2 015 - 2 024
- [11] Guo S, Xiong H, Zheng X, et al. Activity Recognition and Semantic Description for Indoor Mobile Localization[J]. *Sensors*, 2017, 17(3):1-24
- [12] Liu Y, Cheng D, Pei T, et al. Inferring Gender and Age of Customers in Shopping Malls Via Indoor Positioning Data[J]. *Environment and Planning B: Urban Analytics and City Science*, 2020, 47(9): 1 672-1 689
- [13] Wang Peixiao, Wang Haibo, Fu Mengying, et al. Research on Semantic Location Prediction of Indoor Users[J]. *Journal of Geo-Information Science*, 2018, 20(12): 1 689-1 698(王培晓, 王海波, 傅梦颖, 等. 室内用户语义位置预测研究[J]. 地球信息科学学报, 2018, 20(12): 1 689-1 698)
- [14] Li H, Lu H, Shou L, et al. In Search of Indoor Dense Regions: An Approach Using Indoor Positioning Data[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30:1 481-1 495
- [15] Yang Xiping, Fang Zhixiang, Zhao Zhiyuan, et al. Analyzing Space-Time Variation of Urban Human Stay Using Kernel Density Estimation by Considering Spatial Distribution of Mobile Phone Towers[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(1): 49-55(杨喜平, 方志祥, 赵志远, 等. 顾及手机基站分布的核密度估计城市人群时空停留分布[J]. 武汉大学学报·信息科学版, 2017, 42(1): 49-55)
- [16] Li X, Han J, Lee J G, et al. Traffic Density-Based Discovery of Hot Routes in Road Networks: International Symposium on Spatial and Temporal Databases[C]. Berlin, Heidelberg: Springer, 2007
- [17] Chui C K, Chen G. Kalman Filtering with Real-Time Applications[J]. *Applied Optics*, 1987, 28(3): 217-228
- [18] Zheng Y, Zhou X. Computing with Spatial Trajectories[M]. New York:Springer, 2011
- [19] Ankerst M M, Breunig M, Kriegel H P, et al. OPTICS: Ordering Points to Identify the Clustering Structure[J]. *Sigmod Record*, 1999, 28:49-60
- [20] Agrawal K P, Garg S, Sharma S, et al. Development and Validation of OPTICS Based Spatiotemporal Clustering Technique[J]. *Information Sciences*, 2016, 369:388-401

A Method for Identifying Spatial and Temporal Aggregation Area of Indoor Crowd

WANG Peixiao¹ WU Sheng¹ ZHANG Hengcai^{2,3} LU Feng^{2,3} WANG Hong'en⁴

¹ The Academy of Digital China, Fuzhou University, Fuzhou 350002, China

² State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

³ Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou 350002, China

⁴ College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

Abstract: **Objectives:** Compared with the outdoor space, the indoor three-dimensional space structure is complex, and indoor crowd gathering is more likely to lead to safety accidents. As the development of indoor positioning technology, it is possible to collect indoor trajectory, which also provides an important data source for the identification of indoor crowd gathering area. **Methods:** We propose a novel method called IndoorSRC (indoor simplification reconstruction cluster) to detect indoor crowd gathering areas. Firstly, a new indoor user trajectory simplification algorithm, indoor spatial-temporal agglomerative nesting (Indoor-STAGNES), is designed to identify indoor user stay points and simplify indoor user's trajectory. Then, an indoor trajectory reconstruction method based on Kalman filter is constructed to realize the alignment and resampling of indoor user trajectories. Finally, a new indoor space-time density clustering algorithm: Indoor spatial-temporal ordering points to identify the clustering structure (Indoor-STOPTICS) is proposed to find indoor three-dimensional space-time crowd gathering area. **Results:** The real shopping mall indoor trajectory data are used for experimental analysis. The experimental results show that: (1) The gathering areas in the mall are mostly concentrated in the noon time and are mainly located in the dining area. (2) Compared with the traditional outdoor identification method, the error of recognition can be reduced by 23.7% in the case of IndoorSRC with little difference in running time. **Conclusions:** IndoorSRC can provide an effective supplement to the indoor crowd gathering area identification, and provide technical support for early warning and emergency rescue of indoor safety accidents.

Key words: crowd gathering area identification; indoor trajectory; Kalman filtering; Indoor-STAGNES; Indoor-STOPTICS

First author: WANG Peixiao, postgraduate, specializes in geographic information services and spatiotemporal data mining. E-mail: peixiao-wang@whu.edu.cn

Corresponding author: ZHANG Hengcai, PhD, assistant professor. E-mail: zhanghc@reis.ac.cn

Foundation support: The National Natural Science Foundation of China (41771436, 41701521, 41771476); the National Key Research and Development Program of China (2016YFB0502104, 2017YFB0503500); Digital Fujian Program ([2016]23).

引文格式: WANG Peixiao, WU Sheng, ZHANG Hengcai, et al. A Method for Identifying Spatial and Temporal Aggregation Area of Indoor Crowd[J]. Geomatics and Information Science of Wuhan University, 2021, 46(5): 790-798. DOI: 10.13203/j.whugis.20190228 (王培晓, 吴升, 张恒才, 等. 一种室内人群时空聚集区域识别方法[J]. 武汉大学学报·信息科学版, 2021, 46(5): 790-798. DOI: 10.13203/j.whugis.20190228)